

DOCUMENT RESUME

ED 272 394

SE 046 889

AUTHOR Blosser, Patricia E., Ed.; Helgeson, Stanley L., Ed.

TITLE Investigations in Science Education. Volume 12, Number 4.

INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.

PUB DATE 86

NOTE 75p.

AVAILABLE FROM Information Reference Center (SMEAC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (U.S. subscription \$8.00; single copy \$2.75).

PUB TYPE Collected Works - Serials (022) -- Reports - Research/Technical (143) -- Information Analyses (070)

JOURNAL CIT Investigations in Science Education; v12 n4 1986

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS *Academic Achievement; Chemistry; *Cognitive Processes; College Science; Elementary Secondary Education; Genetics; Higher Education; *Learning Strategies; *Learning Theories; Logical Thinking; Models; Perception; Physical Sciences; Science Education; *Science Instruction; Secondary School Science; Self Concept; Sex Differences; *Student Attitudes

IDENTIFIERS *Science Education Research

ABSTRACT

Abstracts and abstractors' critiques of eight science education research studies and two responses to critiques are presented in this journal issue. Each of the studies reviewed focused on a different research theme. Topics addressed include: (1) validation of a learning hierarchy for the mole concept; (2) transfer of learning with emphasis on models and procedures; (3) construct validity of the Longeot test of logical thinking; (4) proportional, probabilistic, and correlational reasoning; (5) relationships between self-concept and achievement of students in a college genetics course; (6) language preferences of freshmen college chemistry students; (7) student perceptions of science teachers, classes, and course content; and (8) effect of teacher gender on secondary students' perception of their science classroom learning environment. Responses to critiques on the studies of student perceptions and the effects of teacher gender conclude this issue. (ML)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED272394

SCIENCE
EDUCATION
INFORMATION
REPORT

SE 046 889



SCIENCE EDUCATION INFORMATION REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

Points of view or opinions stated in this docu-
ment do not necessarily represent official
DERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Robert H. Howe

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 12, Number 4, 1986

THE ERIC SCIENCE, MATHEMATICS AND
ENVIRONMENTAL EDUCATION CLEARINGHOUSE
in cooperation with
Center for Science and Mathematics Education
The Ohio State University

INVESTIGATIONS IN SCIENCE EDUCATION

Editor

Patricia E. Blosser
The Ohio State University

Associate Editor

Stanley L. Helgeson
The Ohio State University

Advisory Board

Susan M. Johnson (1988)
Ball State University

Glen S. Aikenhead (1986)
University of Saskatchewan

John R. Staver (1986)
University of Illinois
at Chicago Circle

David R. Stronck (1987)
California State University-
Hayward

Dorothy Gabel (1987)
Indiana University

National Association for Research in Science Teaching

 Clearinghouse for Science, Mathematics, and
Environmental Education

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 12, Number 4, 1986

Published Quarterly by

The Center for Science and Mathematics Education
College of Education
The Ohio State University
1945 North High Street
Columbus, OH 43210

Subscription Price: \$8.00 per year. Single Copy Price: \$2.75.
Add \$1.00 for Canadian mailings and \$3.00 for foreign mailings.

Volume 12, Number 4, 1986

NOTES FROM THE EDITOR	iii
MISCELLANEOUS STUDIES	1
Griffiths, Alan K.; Heidi Kass; and Alan G. Cornish. "Validation of a Learning Hierarchy for the Mole Concept." <u>Journal of Research in Science Teaching</u> , 20 (7): 639-654, 1983. Abstracted by KEITH S. ANLIKER and JEFFREY R. PRIBYL . .	3
Wollman, Warren. "Models and Procedures: A Classroom Study of Teaching for Transfer." <u>School Science and Mathematics</u> , 83 (2): 122-132, 1983. Abstracted by FRANK A. SMITH, JR.	10
Ahlawat, K. S. and V. Y. Billeh. "The Factor Structure of the Longeot Test: A Measure of Logical Thinking." <u>Journal of Research in Science Teaching</u> , 19 (8): 647-658, 1982. Abstracted by RICHARD DUSCHL	19
Karplus, Robert et al. "Intellectual Development Beyond Elementary School VIII: Proportional, Probabilistic, and Correlational Reasoning." <u>School Science and Mathematics</u> , 80 (8): 673-683, 1980. Abstracted by UBIRATAN D'AMBROSIO	29
Cannon, Roger K. and Ronald D. Simpson. "Relationships Between Self-Concept and Achievement in a College Genetics Course." <u>Journal of Research in Science Teaching</u> , 17 (6): 559-569, 1980. Abstracted by JOEL J. MINTZES	35

Douglas M. Hill, Sheldon R. Baker, Lawrence H. Talley, Malcolm D. Hobday. "Language Preferences of Freshman Chemistry Students: An Exploratory Study." <u>Journal of Research in Science Teaching</u> , 17 (6): 571-576, 1980.	
Abstracted by ELIZABETH KEAN	41
Yager, Robert E. and Ronald J. Bonnstetter. "Student Perceptions of Science Teachers, Classes and Course Content." <u>School Science and Mathematics</u> , 84 (5): 406-414, 1984.	
Abstracted by EUGENE L. CHIAPPETTA	47
Lawrenz, F. P. and W. W. Welch. "Student Perceptions of Science Classes Taught by Males and Females." <u>Journal of Research in Science Teaching</u> , 20 (7): 655-662, 1983.	
Abstracted by JOHN EDWARDS	53
RESPONSES TO CRITIQUES	63
Response by Yager to Chiappetta's Critique	65
Response by Lawrenz and Welch to Edward's Critique	68

NOTES FROM THE EDITOR:

Issue four of Volume 12 of ISE contains critiques of eight different science education research reports and two responses to critiques. The studies are an assorted lot, each focused on a different research topic. Griffiths, Kass, and Cornish reported on their attempts to validate a learning hierarchy for the mole concept. Wollman focused on the transfer of learning with emphasis on models and procedures. Ahlawat and Billeh examined the construct validity of the Longeot test through the use of factor analysis. Karplus and others examined the development of formal reasoning. Cannon and Simpson studied relationships between self-concept and achievement of students in a college genetics course. Hill, Baker, Talley and Hobday explored the language preferences of freshmen college chemistry students. Yager and Bonnstetter reported on student perceptions of science teachers, classes and course content. Lawrenz and Welch investigated the effect of teacher gender on secondary students' perceptions of their science classroom learning environments. In the "response to critiques" section, Yager has provided some rebuttal to Chiappetta's remarks and Lawrenz and Welch have responded to concerns raised by Edwards in his review of their study.

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

MISCELLANEOUS STUDIES

Griffiths, Alan K.; Heidi Kass; and Alan G. Cornish. "Validation of a Learning Hierarchy for the Mole Concept." Journal of Research in Science Teaching, 20 (7): 639-654, 1983.

Descriptors--*Chemistry; Grade 10; High Schools; *Learning Processes; *Psychometrics; Science Education; Scientific Concepts; *Secondary School Science; *Skill Development; *Transfer of Training

Expanded abstract and analysis prepared especially for I.S.E. by Keith S. Anliker and Jeffrey R. Pribyl, Purdue University; Ross Latham, Adrian College.

Purpose

The purposes of this study were:

1. to compare three methods for validating learning hierarchies using one data set.
2. to develop a valid learning hierarchy for the mole concept.

Rationale

The theoretical basis for the development of task analysis-generated learning hierarchies is Gagne's cumulative model of learning. This model suggests that solution of a new task requires that the learner recall or learn related subordinate intellectual skills. Because of this, identification of valid learning hierarchies derived from task analysis can be useful when planning remedial and nonremedial instruction.

Previously, learning hierarchies have been validated using the "test of inclusion" (White and Clark, 1973), a probabilistic scaling method (Dayton and Mcready, 1976) or the "ordering-theoretic method" (Airasian and Bart, 1975). However, the results of these three popular methods have not been compared when applied to the same data set. It is this comparison, with emphasis on the strengths of each method and the potential use of each one in the overall process of hierarchial validation, that is the major concern of this paper.

The mole concept is a key component in introductory chemistry courses and is, therefore, worthy of indepth hierarchial analysis and subsequent validation.

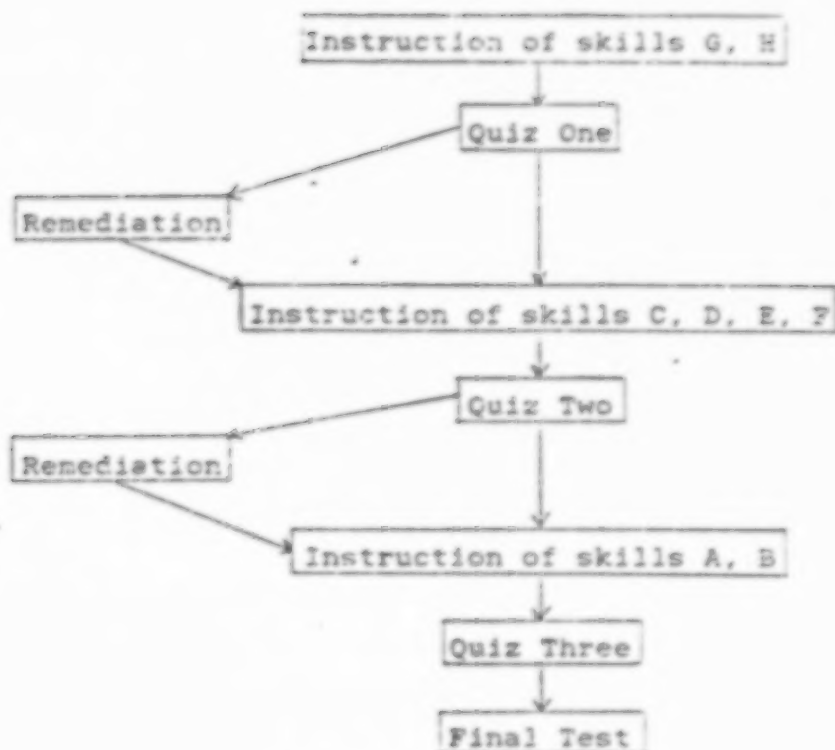
Research Design and Procedure

The sample consisted of 269 grade 10 students taking introductory chemistry courses at four high schools in the Calgary, Canada school system. Eight teachers using the Chem Study materials were involved in the study. The remedial subgroup consisted of four randomly selected classes (109 students) out of the eleven classes used in the study.

The hypothesized learning hierarchy was comprised of the following eight skills identified by the authors:

- "(A) Calculate the masses of different elements or compounds containing the same or proportional numbers of atoms or molecules.
- (B) Convert the mass of an element or compound to the number of atoms or molecules, and vice versa.
- (C) Determine the relative number of atoms or molecules present in given mole quantities of elements or compounds.
- (D) Convert a given number of moles of an element or compound to the number of atoms or molecules present, and vice versa.
- (E) Calculate the mass of an element or compound containing the same number of moles as a given mass of another element or compound.
- (F) Convert a given mass of an element or compound to the number of moles represented, and vice versa.
- (G) Apply the definition of mole as it relates to the Avogadro number of atoms or molecules and to the molar mass of an element or compound.
- (H) Identify and apply the definition of molar mass as a ratio."

These skills were the foundation for the instructional sequence outlined in Figure 1. All quiz and test items were keyed to specific skills in the hypothesized hierarchy. Compilation of student responses to these items generated a single data set for analysis by the methods of White and Clark (1973), Dayton and Macready (1976), and Airarisian and Bart (1975).



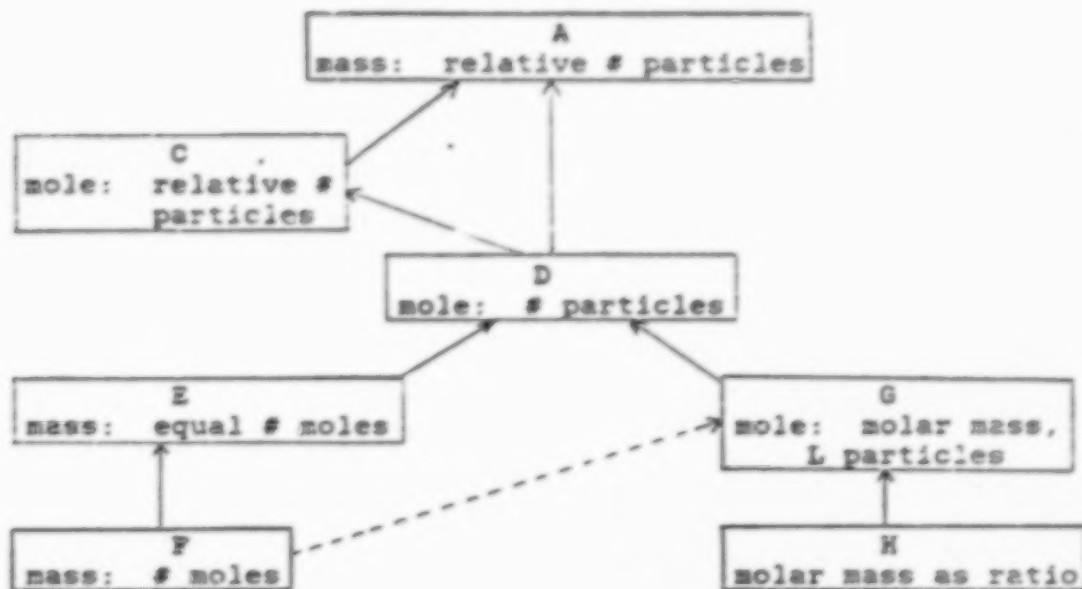
* Figure 1. Instructional Sequence

* Modification of Figure 2 in original paper.

Findings

The investigators found that the originally hypothesized mole hierarchy was inconsistent with the results of the modified White and Clark test (Griffiths and Cornish, 1979) and the ordering-theoretic method. However, those two tests produced a similar set of hierarchical relationships that could be used to construct new hierarchies. The potential hierarchies thus generated were analyzed by the method of Dayton and Macready (1976) to yield a more complex hierarchy. This hierarchy (Figure 2) was identified as the psychometrically validated hierarchy.

The authors did not find significant differences when the scores of remedial and nonremedial groups were compared. However, those students who failed to exhibit a particular subordinate skill on a quiz and later mastered the skill did show significantly greater achievement of related superordinate skills on the final exam.



* Figure 2. Validated Mole Hierarchy

* Combination of Figure 1 and Figure 3(10) in original paper.

Interpretations

The authors drew the following conclusions from the study:

1. The flexibility of the White and Clark method and the ordering-theoretic method suggests their usefulness in preliminary identification of alternative hierarchies that can form the basis for application of the Dayton and Macready method.
2. The Dayton and Macready method is the superior method of validation because of its ability to accomodate disjunctive connections within a hierarchy while also offering a direct test between different hierarchies.
3. A number of skills have been identified, each of which is a necessary prerequisite to the learning of a key skill in the mole concept.
4. There is a need among students to develop the ability to deal with actual numbers of particles on the way to developing the ability to deal with relative numbers of particles.

ABSTRACTOR'S ANALYSIS

The article compares and contrasts the results from the application of different statistical methods for hierarchical validation, but it also makes suggestions for the integrative use of these methods in future work. This emphasis, in fact, is the focus of the study, and these suggestions are certainly the greatest contributions made by this report. It is with this in mind that we suggest a more appropriate (and less misleading) title for the article: "A Comparison of Three Validation Methods for Learning Hierarchies." The title used by the authors leads the reader to expect a definitive study on the teaching and learning of the mole concept, when in reality, the authors' selection of the mole concept is unimportant to the study.

Various questions can be raised concerning the research design and the procedures employed by the authors:

- 1) What was the time frame for the study? How long was each instructional period? What was the length of time between the periods? This rather conspicuous absence of information on the time frame is especially distressing.
- 2) What steps, if any, were taken to control for differences between teachers, classes, and schools? It seems likely that with eight teachers involved in the instructional phase, some sort of monitoring or standardization would be necessary.
- 3) What exactly were the "remedial units"? Were the units diagnostic and instructional in nature, or were they simply drill and practice?

The written report was inadequate in its presentation of data and data analysis. For example, the writers state that "the modified White and Clark test and the ordering-theoretic methods produced similar results," yet the results of the ordering-theoretic method are not presented. ERIC document ED 167 625 (paper presented by Griffiths and Cornish at the Annual Meeting of the American Educational Research Association, March, 1978) gives this information in tabulated form, and in general offers a more complete presentation of the data and data analysis than that given in the JRST article.

Also absent from the written report are examples of the manipulations required by the eight skills and representative test items that were used to measure mastery of these skills. The verbal descriptions of these were not sufficient to clearly define the nature of the tasks required or the expected student responses.

A few changes in the organization of the paper would make the presentation more easily understood. The sample section of the paper makes no mention of the remedial subsample--this information is inappropriately located near the end of the results section. Also found in the results sections is a statement of "the problems of applying the White and Clark test" This material should have been placed with the authors' other evaluative remarks in the discussion portion of the paper.

By comparing in great detail the methods of White and Clark, Dayton and Macready, and Airasian and Bart, the study seems to have filled an important gap in the literature. However, the general reader or the researcher would be well advised to seek out the previously mentioned ERIC document which thoroughly displays the data and describes the data analysis for the study. It appears that the authors chose to include some material about the mole concept in the JRST report at the expense of some important details. It is the abstractors' opinion that this material on the mole concept is not particularly descriptive or prescriptive in nature. In fact, its presence seems only to detract from the authors' real purposes, which we perceive to be the comparative and evaluative aspects of the study.

REFERENCES

- Airasian, P. W., and W. M. Bart. "Validating a Priori Instructional Hierarchies." Journal of Educational Measurement, 12 (2): 163-173, 1975.
- Dayton, D. M. and G. B. Macready. "A Probabilistic Model for Validation of Behavioral Hierarchies." Psychometrika, 41 (2): 189-204, 1976.
- Griffiths, A. K. and A. G. Cornish. "An Analysis and Comparison of Three Recent Methods of Validating Learning Hierarchies." St. John's, Newfoundland, Canada: Memorial University, 1979. ED 167 625.
- White, R. T. and R. C. Clark. "A Test of Inclusion Which Allows Errors of Measurement." Psychometrika, 38: 77-86, 1973.

Wollman, Warren. "Models and Procedures: A Classroom Study of Teaching for Transfer." School Science and Mathematics, 83 (2): 122-132, 1983.

Descriptors--Elementary Secondary Education; Laboratory Training; *Mathematical Applications; *Problem Solving; Science Curriculum; *Science Education; *Science Instruction; *Teaching Methods; *Transfer of Training

Expanded abstract and analysis prepared especially for I.S.E. by Frank A. Smith, Jr., West Chester University.

Purpose

Two questions involving the transfer of learning were asked in this study:

Primary Question: Can "...students substantially and quickly benefit from a teaching strategy that was explicitly designed to promote transfer by providing a physical model and a procedure for applying it to transfer (novel) tasks?"

Secondary Question: Whether students "...in the absence of the model plus procedure students would identify a task as requiring transfer of prior knowledge or whether students would fail to recognize the task as such."

The following null hypothesis was tested:

Hypothesis: There will be no difference between the ratio of the number of students who can successfully complete a transfer task immediately after being exposed to a brief training session, to the total number of students in the group, and this same ratio for a similar group who have not been exposed to the training session.

Rationale

Many science education programs are based upon the assumption that what is formally taught in the classroom or laboratory will transfer to new situations both in school and out of school.

Teachers are encouraged to teach for transfer. An important question for the classroom teacher is, how? This study describes a classroom strategy which could be useful in increasing the transfer of learning. Students first learn in the laboratory a rule for a particular physical system and then are asked questions later about a similar physical system that can be answered by applying the same rule. An experimental group is given a brief training session, designed to promote transfer of learning, immediately prior to performing the transfer task.

This research is related to other studies that attempt to foster transfer of learning (Resnick and Ford, 1979a, 1979b), to studies on the assimilation of a novel task to a previous task (Wollman and Lawson, 1977, 1978), and to studies of general problem solving (Tuma and Reif, 1980).

Research Design and Procedure

The sample used in this study consisted of two classes of a college physical science course. In each class there were 17-18 sophomore or junior elementary education majors. Most had poor science and mathematics backgrounds. The first training session involved balance beam tasks. Two weeks later a second training session involving pendulum tasks was performed. Tables 1 and 2 below show the training sequences and the ratios of the number of students to correctly complete a transfer task to the total number of students in a class.

Table 1
Training Sequence and Class Results For Balance Beam Tasks

Class	Balance-(1 wk)-Lever Transfer-(1 wk)-Lid Transfer Task	Beam Task	Task
AM	Learn	no training 3/17	training 11/17
PM	Learn	training 11/17	no training 11/18

Table 2
Training Sequence and Class Results for Pendulum Tasks

Class	Pendulum Task-(1 wk.)-Meterstick-(1 wk.)-Can Transfer Task	Transfer Task	Task
AM	Learn	training 10/17	no training 10/17
PM	Learn	no training 2/18	training 12/18

BEST COPY AVAILABLE

For the balance beam tasks both classes first studied a balance beam in the laboratory in which they varied the weights and their locations and devised a rule for equilibrium. They also did homework on the balance beam which was gone over in class. One week later, each class took an examination in which three of the questions were related to the study. The first two questions constituted a review of the balance beam. The third question was the lever transfer task. Part of this question asked whether the students thought the lever task was something new or related to something previously studied and, if so, what. After this question was answered, both classes were told that the lever task was a balance beam task "in disguise." It was at this point that one of the classes was given a 10 minute training session on how to remove the disguise. Both classes then attempted to answer the transfer question. A similar procedure was followed one week later for the lid transfer task with the other class now receiving the training session. Several weeks later, the procedure was repeated for the pendulum tasks with the order in which the classes received the training session reversed.

The training sessions involved leading the class to develop a general model of the balance beam or pendulum, the essential features of which could be transferred to similar but novel systems. The training methods are outlined in detail in the article.

The transfer tasks each consisted of a written description of a physical system accompanied by a diagram. Each of the questions is given in the article. The diagrams are omitted.

Findings

Results for the four transfer tasks are shown in Tables 1 and 2. From Table 1 it can be seen that only 3 of 17 untrained students were able to solve the lever transfer task whereas 11 of the 17 trained students were able to solve the task. One week later, 11 of the 17 AM students who now received the training could solve the lid task,

indicating that the training had an effect. Also, 11 of 16 of the PM students, who received no training at this time, could solve the lid task, indicating that their original training had persisted for at least one week. All differences between ratios due to the training session yielded x^2 values significant at $P < .025$.

Table 2 shows the results for the same procedure for pendulum tasks in which the AM class received the training before the PM class. The results indicate that class differences were not responsible for differences in class scores. Also, it can be seen that the training received by the PM class on the balance beam tasks apparently did not transfer to the meterstick task, since only 2 of 18 students could solve the meterstick task.

Students in both classes did uniformly well on the first two questions of the examination relating to the original balance beam task and pendulum task. On the questions regarding the similarity of the transfer tasks to the original tasks, 25 percent of the students recognized that the lever tasks were related to the balance beam task while all of the students recognized that the pendulum transfer tasks were related to the original pendulum task.

Interpretations

The principal conclusions reached by the study were:

1. About half the students could acquire in a very brief time a procedure for transferring balance beam and pendulum knowledge to novel transfer tasks.
2. This learning was retained for at least one week.
3. Without training, most students failed the transfer tasks and failed to recognize that the lever task was a balance beam task.

Two possible interpretations of conclusion 3 were suggested:

1. Students failed to construct a transferable model of the balance beam.
2. Students may have constructed a model of the balance beam but failed to retrieve it because they were searching their memories for something which more closely looked like the lever system.

The author cautioned that nothing in the study indicated that the students acquired the ability to transfer knowledge on their own.

Some implications of the study suggested were:

1. None of the students in this study could recall ever being taught how to apply knowledge in general. Science and mathematics instruction should emphasize the process of achievement as well as what has been achieved.
2. Research is needed to determine how we can teach students to retrieve and evaluate models for use in transfer tasks.
3. The present research is not supportive or critical of Piaget's position (on assimilation) but co-exists with it.

ABTRACTOR'S ANALYSIS

This study essentially asks: will training in how to solve a particular class of problems result in gains in students' ability to solve one of these problems? It is not surprising that the answer found is yes. However, the main contribution of this study is that the training session described is:

1. Very brief (10 minutes)
2. Of an instructional nature that can be employed by a classroom teacher with a large group of students or with individual students.
3. The training is persistent for at least a week for a particular class of problems.

As such, the training procedure can be recommended as a practical and useful teaching strategy for science and mathematics classroom teachers. The descriptions of the training methods given in the study for both the balance beam tasks and the pendulum tasks are detailed enough to be adaptable by teachers from elementary to the college level.

The main threat to the validity of the study comes from the evaluative instrument. The instrument consisted of three questions that were part of a larger examination. The third of these questions was the transfer task and it was the answer to this question that was used to test the research hypothesis. The correct answer to the lever task and the lid task are numerical answers in pounds. The meterstick task is answered by selecting from two choices and explaining the choice. The leaky can task is answered by selecting from three choices and explaining the choice. It is stated that, "Answers were scored right or wrong with no penalty for computation errors or misplacement of the center of mass (for which the students had little practice)." This statement on scoring raises the following questions:

1. How was the explanation considered in scoring the pendulum tasks? If the students' choice of alternative answers is correct and the explanation wrong, is the task scored right or wrong? There seems to be a subjective element in the scoring.
2. How is the phrase "...no penalty for computation errors or misplacement of the center of mass (for which students had little practice)" to be interpreted? It is only the balance beam tasks that require any actual computation, and they ask for only a numerical answer and no explanation.

Other questions arise about the validity of the transfer tasks themselves, such as:

1. The students can find a solution to the lever task as described only if it is assumed that the lever is massless (and hence its unknown mass and the placement of the center of mass can be ignored). If this task described a real lever, its mass and center of mass must be considered. If the students are to ignore the mass of this lever, why isn't this made explicit in the statement of the task? Also, the question for this task is stated as, "How many pounds must be applied at the other end of the pole in order to lift the heavy rock until the pole is horizontal?" The word "until" could be interpreted by a novice problem solver to imply a time sequence requiring some cumulative number of pounds as an answer.

2. The students can obtain a solution to the lid task only if the weight of the lid and the correct placement of the center of mass of the lid are used. The author states elsewhere that the students had little practice in doing this.
3. The pendulum task on the leaky can contains the statement, "Assume that there are no friction effects which slow down the can." In the small angle approximation for a simple pendulum, the frequency of the pendulum is independent of amplitude and, therefore, even if there is friction and the can slows down (its linear velocity), the frequency will not be affected. All students had determined through laboratory work that "...the period of a pendulum depends mainly on the string length and the bob length." They did not find that the period was affected by the amplitude of swing. Why call their attention to it in this task? And if the author feels that it is necessary to call their attention to friction on the can task, why not on the meterstick task where friction may be even greater?

The author assumes that the superior performances on the transfer tasks after the training sessions were due to the students' development of a transferable "model." It should be pointed out that the training session also provided the students with an enhanced vocabulary (both verbal and pictorial) which could affect their performance on the transfer tasks. If one accepts the result that the training session had an effect, it is still not clear what aspect of the training session is responsible for the effect.

In summary, this study describes a training method that may be effectively used by classroom teachers to increase transfer of learning. The validity of the study is threatened by a vagueness in the scoring of the transfer tasks and by the wording of the transfer tasks themselves. Replication of the study by another investigator, based upon the written report, is not possible. Generalization of the results of the study is questionable.

The author may have developed an effective technique for the transfer of learning. It is important enough to repeat the experiment with the scoring procedure and transfer tasks clarified.

REFERENCES

- Resnick, L. B. and W. W. Ford. The Psychology of Mathematics for Instruction, Hillsdale, New Jersey: Lawrence Erlbaum, 1981
- Tuma, D. T. and F. Reif. Problem Solving and Education, Hillsdale, New Jersey: Lawrence Erlbaum, 1980.
- Wollman, W. and A. E. Lawson. "Teaching the Procedure of Controlled Experimentation: A Piagetian Approach," Science Education, 61 (1): 51-70, 1977.
- Wollman, W. and A. E. Lawson. "The Influence of Instruction on Proportional Reasoning in Seventh Graders." Journal of Research in Science Teaching, 15: 227-232, 1978.

Ahlawat, K. S. and V. Y. Billeh. "The Factor Structure of the Longeot Test: A Measure of Logical Thinking." Journal of Research in Science Teaching, 19 (8): 647-658, 1982.

Descriptors--Abstract Reasoning, *Cognitive Development; Cognitive Processes, Cognitive Measurement; Elementary Secondary Education; Elementary School Science, *Factor Analysis; *Logical Thinking; Science Education; Secondary School Science; Tests

Expanded abstract and analysis prepared especially for I.S.E. by Richard Duschl, University of Houston.

Purpose

The intent of the research is to examine the construct validity of the Longeot test for the purpose of testing the "unity hypothesis" presently being debated in the literature. If the factor analysis of the Longeot Test reveals clustering around a single factor, then the "unity hypothesis" is valid. If the factor analysis reveals clustering around two or more factors, then the "unity hypothesis" is invalid.

Rationale

For the past quarter of a century, Piagetian developmental psychology had been registering an increasingly predominant influence upon many branches of educational psychology. Piagetian stages of development of formal operational thinking imply that the construct "logical thinking" is a unitary trait. That is, attainment of stages of cognitive development also assumes the attainment of a set of propositional logic skills. Recently, the "unity" aspects of the "schemas" underlying the structure of logical thinking has become a controversial issue. For example, Lawson, Karplus, and Adi (1978) conclude on the basis of their findings "that the propositional logic and the formal schemata are not part of the same structural unity of mental operations proposed by Piaget" (p. 473). On the other hand, Shayer (1979) found formal operational thinking to be a unitary construct.

To understand the nature of the psychological construct, it is of vital importance to have some valid measures of the construct. The investigation of the construct, logical thinking, obligated the researchers in this field to measure formal operational thinking not only more objectively and parsimoniously but also on a much larger scale, in order to classify students in various stages of cognitive development to employ instructional methods and materials most appropriate to each stage.

The response has been a number of tests of logical thinking (Bart, 1972; Longeot, 1962; Milakofsky and Patterson, 1977; Raven, 1973; Tisher, 1971; Lawson, 1978; Rowell and Hoffman, 1975; and Shayer and Wharry, 1974) which fall into two groups; pure pencil and paper measures and hybrid measures which combine paper and pencil measures with clinical interviews. The Longeot test was selected for the investigation for the following reasons:

- a) It was developed keeping the real spirit of Piagetian concept of developmental stages of logical thinking.
- b) It has stood the test of time and proved useful in a variety of research studies in several countries (Blake, 1980; Lawson and Blake, 1976; Longeot, 1962, 1965).
- c) It employs several items to measure each one of the four operational concepts so as to provide a reliable assessment of each concept.
- d) It is easy to administer and does not require more than a normal instructional period of time.

The Longeot test contains four distinct parts: part 1 includes five class inclusion items; part 2, six propositional logic items; part 3, nine proportional reasoning items; and part 4, eight combinatorial analysis items. Specific research questions are:

- 1) determine the factorial structure underlying the construct logical operational thinking as measured by the Arabic Version of the Longeot test in eleventh grade science students in Jordan;
- 2) provide a meaningful interpretation to the orthogonal dimensions uncovered by factor analysis;

- 3) establish certain psychometric properties of the test for the Jordanian high school science students;
- 4) provide fresh evidence for the validation process of the Piagetian construct of logical thinking.

Research Design and Procedure

The population of the study was twenty-one eleventh grade classes in three public secondary schools. The sample was 209 males and 180 females selected as part of intact classes. The mean age of the students was 17 with a range of 15-20 for the males and 17-19 for the females.

The test was administered to all classes at the beginning of the second semester of the 1979-80 academic year. No differences due to gender or school effects were found from a preliminary analysis of data, hence for all further statistical analysis the data from all classes and schools were merged together.

Intercorrelations of all 28 items on the Longeot test were calculated. The matrix of interitem product moment correlations was then subjected to principal component analysis using Systems Support Programs (IBM). On the basis of the scree test (Cattell, 1966) it was decided to conduct four-, five-, and six-factor solutions via varimax rotations. (Note: Varimax rotations are used in the analysis of multi-variate data utilizing eigenvalues.) Each of the factor analysis solutions obtained simple structures.

Findings

A comparison of the composition of the factors in four-, five-, and six-factor varimax rotated solutions reveals:

- 1) simple structure was attained in all the three instances.

- 2) items 1-5, class inclusion items, loaded on the same factor in all three solutions.
- 3) items 21-25, all combinatorial thinking items, defined the same factor in all three solutions.
- 4) items 26-28, also combinatorial thinking items, did not stick with items 21-25 but defined an independent factor.
- 5) in the four-factor solution, items 6 and 11 (propositional thinking) and 12-17 (proportional thinking) loaded on the same factor, whereas in both five- and six-factors rotations, items 15-20 (propositional thinking) invariably defined a distinct factor.
- 6) items 6-14 showed erratic behavior distributing loadings over several factors. It was only in the six-factor solution that interpretable behavior existed. Hence, only the six-factor solution is interpreted.

Interpretations

Based on the varimax rotation results of the factor analyses, only the six-factor solution was interpreted. The choice of nomenclature in the interpretation of the factors was merely a matter of taste than of scientific terminology. For the sake of convenience, factors were named after the nature of logical operations required by the contents of the items; Factor (F) 1 - combinatorial thinking, F 2 - comparative proportional thinking, F 3 - permutational thinking, F 4 - class inclusion, F 6 - propositional thinking. F 5 was unnamed due to ambiguities attributed to two test items (8 and 10) with high difficulty and low discrimination.

The results of the study cast some doubt on the assumed unitary nature of the Piagetian construct. The data of the present study have yielded at least six independent factors. On the basis of that, it is concluded that whatever the trait measured by the Longeot test is, it is not unitary in nature. Either the Longeot test is not a valid measure of logical thinking or logical thinking is a multidimensional construct.

Implications for science education suggest that grade 11 students might be employing independent strategies for solving logical problems involving concepts identified in the factors listed above. Therefore, curriculum planning and instructional strategies based on the assumption of logical thinking as a unitary construct may not be appropriate.

ABTRACTOR'S ANALYSIS

Interest among science educators with the application of Piaget's theories of cognition to science instruction has been and continues to be a prevalent theme in science education research. In fact, it is appropriate to take the position that this domain of science education research has come of age, given the number of synthesis articles appearing in the literature addressing Piaget's ideas of formal reasoning and stages of cognitive development; some complimentary (Linn, 1982; Lawson, 1982; Staver, 1984) and some antagonistic (Nagy and Griffiths, 1982; Munby, 1980; Driver, 1978). Essentially a clearly defined scholarly exchange is taking place in the literature, with some authors endorsing the application of Piagetian theories to education and others not endorsing such an application.

The authors of the abstracted article are clearly associated with the group endorsing the application of Piaget's theories to education. Within this group though, there exists another controversy alluded to by the authors, the question of whether or not certain propositional logic skills are strongly correlated with cognitive stages of development; specifically the concrete operational stage and the formal operational stage. The authors cite the work of Lawson, et al (1978) and Shayer (1979) as examples of research falling on opposite sides of the "unity hypothesis" argument. Shayer (1983) in a rejoinder to the abstracted article disagrees, "My own work does not, I think, contradict the Lawson, Karplus, Adi one. In fact I quote both that article and the work of Wason and Johnson-Laird as evidence that Piaget's logical model is invalid (p. 707)."

Shayer's reference to the work of Wason and Johnson-Laird (1972) identifies an aspect of the "unity hypothesis" that Ahlawat and Billeh (1982) did not integrate into their article. Specifically the position that an individual's ability to solve problems in propositional logic is affected by the context in which the problems are drafted. Wason and Johnson-Laird (1972) demonstrated using a sample of postal workers that their abilities to solve logic problems with unfamiliar content was lower than when the same logic problems were given in a familiar context. Bady (1979) has followed this lead for testing students' understanding of the logic of hypothesis testing by providing the tasks in a familiar context. From the examples given in the appendix of the article, it appears the Longeot test is potentially dealing with content unfamiliar to the sample of students taking the exam. In spite of the rationales given for using the Longeot test, the use of a test with items presented in the context of science, may have provided a clearer analysis of the issues being investigated.

The researchers' use of factor analysis to examine the construct validity of the Longeot test is also questioned by Shayer (1983). Accepting the design as legitimate, it was difficult to comprehend why the four- and five-factor rotations were not analyzed or interpreted. Furthermore, the article presented several inconsistencies such as the reporting of items 12-17 as proportional thinking, items 15-20 as propositional thinking, and items 12-14 as probabilistic thinking (p. 653). Another inconsistency is the reporting of item 10 having "a tenuous relationship with item 25." (p. 651). Table II of the article reports no correlation coefficient between 10 and 25, but does report significant relationships between 10 and 12, 16, 17, 24, and 27.

The research does demonstrate and contribute to the need to conduct cross-cultural studies on related populations of students. More comparisons of research results in differing educational, social, and cultural contexts should be carried out to establish a broader set of criteria for reliability and validity. To this point the authors are

to be commended. It seems appropriate, then, to extend to some of the newer instruments testing logical thinking the same set of criteria for the purpose of testing the unity hypothesis.

But the debate about the unity hypothesis will wage on; Zeidler (1985) reports that in a factor analysis of Tobin and Capie's (1981) Test of Logical Thinking (TOLT) the five variables of the test (proportional reasoning, controlling variables, probabilistic reasoning, correlational reasoning, and combinatorial reasoning) do represent a single cognitive construct, namely formal reasoning. Furthermore, Zeidler (1985) reports, "the specificity values for each variable and empirical support for the existence of more rudimentary, singular cognitive structures which Piaget, et al (1977) and Inhelder and Piaget (1964) theorized to develop in parallel stages." (p. 466).

Reasoning has been selected by a blue ribbon panel of educators "as the component of literacy most in need of further study." (Glasman, Koff, and Spiers, 1984, p. 464). A discussion of what reasoning involves and means to different disciplines is addressed in a special issue of Review of Educational Research. The contributors clearly establish that research on reasoning is an on-going and rapidly evolving enterprise with many unsolved problems and unaddressed issues. As was stated above, the scholarly discussion on reasoning and on Piaget's theories of cognitive development is both active and welcomed in the science education research arena, and is a measure of the maturity of the conceptual understandings we have of the problems which face us.

REFERENCES

- Ahlawat, K. and V. Billeh. "The Factor Structure of the Longeot Test: A Measure of Logical Thinking." Journal of Research in Science Teaching, 19 (8): 647-658, 1982.
- Bady, R. "Student's Understanding of the Logic of Hypothesis Testing." Journal of Research in Science Teaching, 16 (1): 61-65, 1979.
- Bart, W. M. "Construction and Validation of Formal Reasoning Instruments." Psychological Reports, 30: 663-670, 1972.
- Blake, A. J. D. "The Predictive Power of Two Written Tests of Piagetian Developmental Level." Journal of Research in Science Teaching, 17 (5): 435-441, 1980.
- Cattell, R. B. "The Meaning and Strategic Use of Factor Analysis." in R. B. Cattell (Ed.), Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, 1966.
- Driver, R. "When is a Stage not a Stage? A Critique of Piaget's Theory of Cognitive Development and its Application to Science Education." Educational Research, 21: 54-61, 1978.
- Glasman, N., R. Koff, and H. Spiers. Preface, Review of Educational Research, 54 (4): 461-472, 1984.
- Inhelder B., and J. Piaget. The Early Growth of Logic in the Child. New York: W. N. Norton, 1964.
- Lawson, A. "The Nature of Advanced Reasoning and Science Instruction." Journal of Research in Science Teaching, 19 (9): 727-742, 1982.
- Lawson, A. E. "The Development and Validation of a Classroom Test of Formal Reasoning." Journal of Research in Science Teaching, 15 (1): 11-24, 1978.
- Lawson, A. E. and A. J. D. Blake. "Concrete and Formal Thinking Abilities in High School Biology Students as Measured by Three Separate Instruments." Journal of Research in Science Teaching, 13 (3): 227-235, 1976.
- Lawson, A. E., R. Karplus, and H. Adi. "The Acquisition of Propositional Logic and Formal Operational Schemata During the Secondary School Years." Journal of Research in Science Teaching, 15 (6): 465-478, 1978.
- Linn, M. "Theoretical and Practical Significance of Formal Reasoning." Journal of Research in Science Teaching, 19 (9): 719-725, 1982.

- Longeot, F. "Un Essai d'Application de la Psychologie Genetique a la Psychologie Differentielle." Bulletin de l'institut National D'Etude, 18: 153-162, 1962.
- Longeot, F. "Analyse Statistique do Trois Tests Genetique Collectifs." Bulletin de l'institut National D'Etude, 20: 219-235, 1965.
- Milakofsky, L. and H. O. Patterson. Chemical Education and Piaget: A New Paper-and-Pencil Inventory to Assess Cognitive Functioning. University Park, PA: Pennsylvania State University Press, 1977.
- Munby, H. "Criticizing the Piagetian Orientation to Change." in H. Munby, G. Orpwood, and T. Russell (eds.) Science Curriculum in a New Light. Toronto: OISE Press, 1980.
- Nagy, P. and A. Griffiths. "Limitations of Recent Research Relating Piaget's Theory to Adolescent Thought." Review of Educational Research, 52 (4): 513-556, 1982.
- Piaget, J., J. Grize, A. Szeminska, and U. Bang. Epistemology and Psychology of Functions. Boston: D. Reidel, 1977.
- Raven, R. J. "The Development of a Test of Piaget's Logical Operations." Science Education, 57 (3): 377-385, 1973.
- Rowell, J. A. and P. J. Hoffman. "Group Tests for Distinguishing Formal From Concrete Thinkers." Journal of Research in Science Teaching, 12 (2): 157-164, 1975.
- Shayer, M. "Comments on the Factor Structure of the Longeot Test: A Measure of Logical Thinking." Journal of Research in Science, 20 (7): 707-708, 1983.
- Shayer, M. "Has Piaget's Construct of Formal Operational Thinking Any Utility?" British Journal of Educational Psychology, 49: 256-276, 1979.
- Shayer, M. and D. Wharry. "Piaget in the Classroom, Part 1: Testing a Whole Class at the Same Time." School Science Review, 55: 447-458, 1974.
- Staver, J. "Research on Formal Reasoning Patterns in Science Education: Some Messages for Science Teachers." School Science and Mathematics, 84 (7): 573-589, 1984.
- Tisher, R. P. "A Piagetian Questionnaire Applied to Pupils in a Secondary School." Child Development, 42: 1633-1636, 1977.
- Tobin, K. and W. Capie. "The Development and Validation of a Group Test of Logical Thinking." Educational and Psychological Measurement, 41: 413-423, 1981.

Wason, P. C. and P. N. Johnson-Laird. Psychology of Reasoning:
Structure and Content. Cambridge: Harvard U. Press, 1972.

Zeidler, D. "Heirarchial Relationships Among Formal Cognitive
Structures and Their Relationship to Principled Moral
Reasoning." Journal of Research in Science Teaching, 22 (5):
461-471, 1985.

Karplus, Robert et al. "Intellectual Development Beyond Elementary School VIII: Proportional, Probabilistic, and Correlational Reasoning." School Science and Mathematics, 80 (8): 673-683, 1980.

Descriptors--*Abstract Reasoning; *Cognitive Development; Elementary Secondary Education; Higher Education; *Intellectual Development; Mathematics Education; Ratios Mathematics; Science Education; *Science Tests

Expanded abstract and analysis prepared especially for I.S.E. by Ubiratan D'Ambrosio, Universidade Estadual de Campinas.

Purpose

To gather evidence showing that curricular separation of studies of mathematics and science may work against the development of formal reasoning.

Rationale

This paper, the eighth of a series, tackles the very serious question of the development of formal reasoning, which has been increasingly questioned in its relation to student success in secondary and college science and mathematics courses. The authors examine the problem using six group-administered tasks which have been considered before and which require basically proportional, probabilistic and correlational reasoning.

Research Design and Procedure

The subjects were 505 students from sixth grade to college sophomores, with ages ranging from 11.5 to 20.0 years, with the majority-middle to upper-middle-class in a suburban community. They were closely distributed among the seven educational levels (sixth grade to college),

with a large predominance of males (291 to 214).

Briefly the six tasks were the following:

1. -Proportions: dealing with hollow cylinders of different diameters which are filled with water up to different heights.
2. -Probability: white and black cubes in different sacks.
3. -Probability: sack with flat pieces of wood of different colors (6 altogether).
4. -Correlations: sets of fat and thin mice with black and white tails in a picture, and the task is to correlate size of mice and color of tails.
5. -Probability: what are the chances of another mouse to be fat?
6. -Correlations: sets of fishes (large and small) with stripes (narrow and wide). Correlation between size of fish and width of stripes.

The responses of all the 505 subjects were categorized according to the following categories, which reflect primarily the explanation by which the students justified their answers. The categories were named in the following way, for each of the six tasks:

Task 1: I (intuitive), A (additive), Tr (transitional), R (ratio).

Task 2: I, AV (absolute value), IC (one comparison), 2C (two comparisons).

Tasks 3 and 5: I, Ap (approximate), Q (quantitative).

Tasks 4 and 6: I, NR (no relationship), TC (two cells), FC (four cells), Co (correlation).

The names adopted for the categories correspond, to a certain extent, to associated behaviors of the subjects when facing the tasks, and are compatible with those introduced by R. Karplus and others.

Findings

The findings tabulated in the paper show basically that for Task I most of the sixth graders used additive procedures (category A), while the majority of upper-middle-class high school students used proportional reasoning (category R). For college students, proportional reasoning was the primary response, while a minority (about 20 percent) responded in categories I or A.

On Task 2, more than half the sixth graders used two comparisons to justify their conclusions, while only 38 percent of these students used absolute value or one comparison. Use of two comparisons was over 50 percent in all grades, increasing from 55 percent for 6th graders to 93 percent for college students.

For Tasks 3 and 5, the gradual increase in the percentage of students who responded in Category Q in both tasks, displaying probabilistic reasoning consistently is noticeable.

For Tasks 4 and 6, there is a marked decline in category I (intuitive) from 65 percent for 6th graders to 14 percent for college students, while correlation appears with "zero" answers for 6th graders up to 50 percent for college students.

Among the conclusions the authors mentioned that the term "relation" is difficult for younger students' explanations on the correlative items.

Interpretations

The authors present an overview of the categories used in the research grouped according to the level of reasoning they reflect. They distinguish between formal level, transitional level, and concrete level of reasoning, and analyse the results in order to group them accordingly. Summarizing, they try to answer the following questions: (1) What categories are required for classifying the subjects' responses on tasks requiring proportional, probabilistic, or correlational reasoning? (2) How effective are these tasks for assessing these aspects

of formal reasoning and (3) What implications for teaching are suggested by the observed distributions of students' responses among the categories required?

Among the interpretations of the authors, we distinguish the suggestion that science and social studies courses should pay more attention to the analysis of data for correlations. Also, they suggest that teachers should be cautious in emphasizing activities heavily dependent on correlational reasoning in junior high level.

As a major conclusion, the authors recommend the close connection between mathematics and science teaching with the aspects of formal reasoning, and indeed advocate the tying together of science and mathematics in schools.

ABTRACTOR'S ANALYSIS

This is a most welcome research. Although somewhat in the general line of Piagetian research, it has very concrete and original aims and the method is potentially very rich. We do hope this line of research finds more space among science and mathematics educators.

It has long since been recognized that the development of formal reasoning can be related to the success in secondary and college science and mathematics courses. Fortunately for science educators! But not much has been done in recognizing the development of steps of formal reasoning among adolescents and young adults. The range 11.5 to 20.0 years is a most fortunate choice by the authors for this research, not only from the view of pure cognitive development, but also from the relationship science-society. Only briefly the authors mention everyday concerns, such as cigarette smoking, car accidents, etc., in relation to science education. Probably this could be pursued further.

The methodology is not surprising. The categories used in the paper make sense and in fact among the results we have a discussion on the choice of categories. Again, this makes this paper a sort of open box for much research in the same framework. Some bias in favor of

probability might be questioned and indeed we might seek an explanation in the context of cultural factors underlying a research project. Which cultural factors have influenced the researchers in selecting the proposed six tasks? In a sense, this remark reflects much of the criticism on Piaget. This does not invalidate the methodology and the results, but it has not been considered in cognitive development. Very briefly, the authors mention social background of the subjects.

The research design is sound and provocative, in the sense that it is simple enough to entice further research in this line. The categories used to classify the responses, which are indeed the key element in the research, reflect a long history of many years of research by the first author and many collaborators. They cover much of what one identifies as logical behavior. The categories reflect much of the biases mentioned above, hence can be modified if one looks for behavior not so heavily relying on proportional, probabilistic or correlational reasoning. But indeed these three kinds of behavior are what the authors had in mind, as it is shown in the title of the paper itself. Hopefully, other papers will be presented on other kinds of behavior than those pursued by the first author and a number of collaborators in the series of papers, in which the present paper is number VIII. Indeed, the three basic questions mentioned in the interpretations section of this abstract may serve as guidelines for future research in this direction.

These three questions which, in a sense, synthesize much of the focus of the research generate many interesting side results. For example, what is known to most science educators comes up again: limitations are due to difficulties of the subjects to comprehend written instructions and to express themselves in writing. On the other hand, it is pointed out the absence of cueing due to body language of the interviewer as one advantage of this kind of test. Also, copying and students helping each other was taken into account.

The paper is written in an equilibrated way, with a sound distribution into sections and easy to understand and interpret tables, six in the total. Surely, many interesting comments on specific responses are available to the authors. Just one of such responses by

a sixth grader is reported on page 681. We would like to see many other examples. Although not important for the present paper, an appendix might suggest to us some alternative patterns of logical behavior which would directly affect the choice of categories, introducing some other categories which would reflect cultural background, as we have mentioned above.

It should be mentioned again, and in fact these remarks particularly pleased the abstractor, the conclusions about integrated curricula for science and mathematics. The authors go beyond explicit conclusions about the close connection of both mathematics and science teaching with the aspects of formal reasoning investigated. They stress the fact that the present separation of studies into mathematics, which in its turn is separated into numerical, algebraic and geometric aspects, and science, which is also separated into different subjects, does not allow the student to proceed to a significant collection and analysis of data related to a real problem. Usually, data are gathered in science courses, but mathematical manipulation is not carried out. While in mathematical courses, data are manipulated, but usually these are meaningless data. The authors are very clear and straight to the point in urging that mathematics and science curricula be tied together. They hint, although less explicitly, on different approaches to the solution of problems which arise from real life situations. Indeed, the solution of problems, so much emphasized nowadays, should not be carried on without the previous step of formulation of the problem, i.e., raise questions and formulate adequate hypotheses. The solution, which results from adequate testing of hypothesis and analysis of conclusions tied together with mathematical manipulation of data, can not be thought independently of formulation. This modelling approach to problem solving seems to be lacking in our school systems and this paper may serve as a preliminary backing for this approach. Surely, we would like to see both integrated science and mathematics curricula and modelling approach to problem solving as areas of research to be pursued as a consequence of the important research reviewed.

Cannon, Roger K. and Ronald D. Simpson. "Relationships Between Self-Concept and Achievement in a College Genetics Course." Journal of Research in Science Teaching, 17 (6): 559-569, 1980.
Descriptors--Academic Achievement; Biological Sciences, *College Science, *Genetics; Higher Education, Science Education; *Self Concept, *Sex Differences

Expanded abstract and analysis prepared especially for I.S.E. by Joel J. Mintzes, The University of North Carolina at Wilmington.

Purpose

This study examined relationships between self-concept and achievement in a college genetics course. In addition, the investigators indicated that they "were interested in seeing if there were significant differences in self-concept and achievement between men and women and between science majors and nonscience majors."

Rationale

A significant body of evidence suggests that self-concept plays a substantial role in human accomplishment (Bloom, 1976; Purkey, 1970). ". . . student failures in the basic school subjects, misdirected motivation, and lack of commitment to develop to their full potential are consequences of their own deficient perceptions of the self and the world."

Design And Procedures

Eighty-six (86) students who enrolled in a "Genetics in Human Affairs" course at North Carolina State University served as subjects. Each student was administered the Bills Index of Adjustment and Values (a global measure of self-concept) and the Brookover Post High School Self-Concept of Ability Scale (a measure of academic self-concept). Relationships among these measures and final course grades were examined by sex and major.

The Bills Index is composed of a set of eight subscales which enabled the investigators to categorize each student by self-concept and perception of others' self-concepts. Thus, students were classified as either ++ (positive "self" and positive "others"), -+, +-, or --. Individuals who fall in the ++ category are said to be secure, cooperative and responsible, and sensitive to the rights and needs of others (Bills, 1975). Those in the -+ category distrust self and yield readily to arbitrary authority. The +- individuals are fundamentally -+ people who defend self against others and thus are basically distrustful yet respectful of constituted authority. In the -- category are people who "have a long list of unhappy times and blame other people for their unhappiness."

The data were analyzed in four ways. First, intercorrelations among the eight subscales of the Bills Index, the Brookover Scale, and achievement were calculated. Second, variation in achievement attributable to sex, major, class, age, Bills and Brookover scores were examined by analyses of variance. Third, the relative importance of each of the variables in predicting achievement was assessed in a stepwise regression analysis. And finally, relationships among self-concept categories (++, -+, +-, --), major, sex, and achievement were explored in a contingency table (cross tabulation) analysis.

Findings

(1) Intercorrelations:

- (a) Significant relationships were found among several of the Bills Index subscale scores, and among these scores and Brookover Scale scores.
- (b) Relationships among Bills Index scores and achievement were nonsignificant for seven of the eight subscales and weak ($r = -.22$, $p < .04$) for the eighth.
- (c) A significant relationship ($r = .53$, $p < .0001$) was found between Brookover Scale scores and achievement.

Cannon, Roger K. and Ronald D. Simpson. "Relationships Between Self-Concept and Achievement in a College Genetics Course." Journal of Research in Science Teaching, 17 (6): 559-569, 1980.

Descriptors--Academic Achievement; Biological Sciences, *College Science, *Genetics; Higher Education, Science Education; *Self Concept, *Sex Differences

Expanded abstract and analysis prepared especially for I.S.E. by Joel J. Mintzes, The University of North Carolina at Wilmington.

Purpose

This study examined relationships between self-concept and achievement in a college genetics course. In addition, the investigators indicated that they "were interested in seeing if there were significant differences in self-concept and achievement between men and women and between science majors and nonscience majors."

Rationale

A significant body of evidence suggests that self-concept plays a substantial role in human accomplishment (Bloom, 1976; Purkey, 1970). ". . . student failures in the basic school subjects, misdirected motivation, and lack of commitment to develop to their full potential are consequences of their own deficient perceptions of the self and the world."

Design and Procedures

Eighty-six (86) students who enrolled in a "Genetics in Human Affairs" course at North Carolina State University served as subjects. Each student was administered the Bills Index of Adjustment and Values (a global measure of self-concept) and the Brookover Post High School Self-Concept of Ability Scale (a measure of academic self-concept). Relationships among these measures and final course grades were examined by sex and major.

The Bills Index is composed of a set of eight subscales which enabled the investigators to categorize each student by self-concept and perception of others' self-concepts. Thus, students were classified as either ++ (positive "self" and positive "others"), +-, or --. Individuals who fall in the ++ category are said to be secure, cooperative and responsible, and sensitive to the rights and needs of others (Bills, 1975). Those in the +- category distrust self and yield readily to arbitrary authority. The -- individuals are fundamentally -- people who defend self against others and thus are basically distrustful yet respectful of constituted authority. In the -- category are people who "have a long list of unhappy times and blame other people for their unhappiness."

The data were analyzed in four ways. First, intercorrelations among the eight subscales of the Bills Index, the Brookover Scale, and achievement were calculated. Second, variation in achievement attributable to sex, major, class, age, Bills and Brookover scores were examined by analyses of variance. Third, the relative importance of each of the variables in predicting achievement was assessed in a stepwise regression analysis. And finally, relationships among self-concept categories (++, +-, --), major, sex, and achievement were explored in a contingency table (cross tabulation) analysis.

Findings

(1) Intercorrelations:

- (a) Significant relationships were found among several of the Bills Index subscale scores, and among these scores and Brookover Scale scores.
- (b) Relationships among Bills Index scores and achievement were nonsignificant for seven of the eight subscales and weak ($r = -.22$, $p < .04$) for the eighth.
- (c) A significant relationship ($r = .53$, $p < .0001$) was found between Brookover Scale scores and achievement.

(2) Analyses of Variance

Of thirteen sources of variance, two explained a significant share of variance in achievement. The significant sources of variance were Brookover Scale scores ($p < .0001$) and college class, i.e. freshman, sophomore, etc. ($p < .012$).

(3) Regression Analysis:

Fourteen variables were added in a stepwise manner to produce a multiple regression model which accounted for 51% of the variance in achievement (e.g., $R = .513$). The Brookover Scale scores were the best single predictors of achievement.

(4) Contingency Table Analysis:

- (a) Differences in achievement among students of the four self-concept categories were not apparent; those in the -- category achieved as well as those in the ++ category.
- (b) No differences were found in self-concept between men and women (overall).
- (c) Among women majoring in nonscience subjects, more fell in the ++ category and fewer in the -- category than in science subjects. Likewise, men majoring in science had more positive self-concepts than did women.
- (d) No significant differences were found in achievement between science majors and nonscience majors.

Interpretations

The authors suggest that "the findings of this study provide further evidence related to the correlation between self-concept and achievement. However, these findings are inconclusive and thus offer topics for further research."

ABTRACTOR'S ANALYSIS

Relationship to Other Studies. The authors identify a number of studies which focus on relationships between self-concept and achievement. More details on specific findings of previous work in science-related disciplines would have been welcome.

Though the authors do cite a number of prominent workers in the field of self-concept (Allport and Odbert, 1936; Bloom, 1976; Brookover et al., 1965; Purkey, 1970; Rogers, 1951), it is difficult to see how this particular study adds significantly to that fund of knowledge. Perhaps what is missing is a coherent theoretical matrix--a conceptual framework which ties these studies together. Such a theory would pinpoint significant variables to study and identify explicit hypotheses worth investigating. In the absence of a theory one is given the impression that researchers are "fishing" for significant effects.

As the authors themselves suggest, the study posed more questions than it resolved. Though several statistically significant effects were found, the most important finding seems to be a single correlation coefficient (that relating Brookover Scale scores and achievement).

Research Design and Written Report. The principal question posed by this study was the effect of self-concept (along with college major and gender) on achievement in a college genetics course. This problem is a significant one and well worth studying; however, several design and analysis problems detracted from the value of this report. One might ask, for example:

- (1) How was achievement measured? It appears that final course grades were used; however, the reader is given no information about the testing instruments (format, reliability, validity). Furthermore, final course grades are a notoriously subjective measure of learning.
- (2) Why were numerical reliability estimates of the self-concept scales omitted? (We are told that the reliability of the adult IAV is "high"). The reader needs this information to properly interpret the findings. If not available from other sources (as with the Brookover Scale), they should be calculated.

- (3) Why were multiple data analyses performed? Often when investigators employ multiple analytic techniques it suggests the possibility of an underlying weakness in the experimental design. For example, in this study the major question might have been addressed in a 2 (sex) x 2 (major) x 4 (self-concept category) factorial design with achievement as a dependent variable. The results would have informed the reader about main effects and, most importantly, about any significant interactions. This type of analysis, however, demands a sufficient number of subjects to fill the available cells.

The written report was comprehensive in its treatment of experimental procedures and results, though it suffered in several respects:

- (1) The statement of purpose ("aims") would have been enhanced by an explicit list of research hypotheses.
- (2) The major conclusions were difficult to disentangle as they were embedded in a somewhat lengthy discussion of instrumentation.
- (3) The report was cluttered with details of instrumentation and a seemingly unnecessary table which identified sample sizes and frequencies of self-concept groups based on different statistical treatments. (The latter could have been included in a previous table.)

Further Research. In reviewing previous studies on the relationship of self-concept and achievement (Brookover *et al.*, 1965; Purkey, 1970) it becomes clear that some of the weaknesses of the present study reflect the embryonic state of research and theory in the field. The authors are correct in their assertion that "one critical area of research is instrument development and validation." It may well be, however, that fundamental inquiry into the nature of "the self" is required before progress on instrumentation and educational applications can be made. Clearly this work does not fall within the province of the science educator but belongs more properly to the field of psychology.

Perhaps another area worthy of further research is the long-term effects of schooling on self-concept. In-depth, longitudinal studies which document the effects on self-concept of success and failure from the early elementary grades through the college years would be very enlightening. This type of work might shed light on the development of the academic self-concept and its subsequent effects on achievement in and attitude toward school.

REFERENCES

- Allport, G. and H. Odbert. "Trait Names: A Psycholexical Study." Psychological Monographs, No. 211, 1936.
- Bills, R. E. A System for Assessing Affectivity. Alabama: University of Alabama Press, 1975.

Douglas M. Hill, Sheldon R. Baker, Lawrence H. Talley, Malcolm D. Hobday. "Language Preferences of Freshman Chemistry Students: An Exploratory Study." Journal of Research in Science Teaching, 17 (6): 571-576, 1980.

Descriptors--Chemistry; *College Science; *College Students; Higher Education; *Language Usage; *Prose; Science Education; *Science Teachers

Expanded abstract and analysis prepared especially for I.S.E. by Elizabeth Kean, University of Wisconsin-Madison.

Purpose

This work has a two-fold purpose:

- 1) to develop a language preference test within the domain of introductory level chemistry content:
- 2) to use the developed language test to obtain tentative answers to the following questions:
 - "Do chemistry instructors have a consistent preference for particular ways of expressing ideas in chemistry?
 - Do chemistry students have a consistent preference for particular ways of expressing ideas in chemistry?
 - Are any such preferences on the part of chemistry instructors shared by the students they teach?
 - Do the preferences of chemistry students differ from those of nonchemistry students?"

Rationale

The authors postulated that college trained chemists would have a stable and measurable preference for certain language patterns when thinking about and expressing ideas in chemistry. Further, the authors stated that chemistry instructors commonly use language differently in different teaching situations (e.g., lectures, tutorials, labs) but that it is possible that freshman students often are not fully aware of language conventions and appropriateness of language to a specific context.

Following researchers such as Heath (1964), they believed a multiple choice format for a language preference test would be appropriate. Many items on the developed test were taken from previous chemistry language work by Shoenfeld (1974).

Research Design

Test Development. Twenty one test items were constructed by the authors. Each item consisted of four statements; each statement described or explained the same situation in a different language style. The language of one of the four statements was "formal technical prose" and was intended to be appropriate for a formal introductory chemistry test situation in which the intended audience for the response was a chemistry instructor. No information was provided about how the alternative statements were constructed.

Test items were of three kinds. Some were concerned with "more technical aspects of scientific communication, e.g., restricted use of ordinary words or phrases such as . . . 'analyses by means of a piece of complex equipment'." Others tested understanding of "more general aspects of scientific communication, e.g., use of past tense, passive voice. . . ." The last category included items which "reflected different levels of understanding of the nature of scientific enquiry, e.g., distinguishing between a system and a model developed to describe that system or between an objective description of the system and teleological or anthropomorphic equivalents." No information was provided about the number of each type of item.

Using the Test. The test was revised "after consultation" with science faculty at one Australian college (Riverina College). The test was then given to five chemistry instructors at that same college and to five college chemistry faculty at an American college (West Liberty State College). Four of the five chemistry faculty at each college chose the same alternative on 20 items (one assumes that these were the "formal technical prose" statements as described above).

These 20 items (face validity assumed) were given to four intact (nonequivalent, nonrandomly chosen) groups of students in a posttest only research design. At Riverina College, one class of chemistry students and one language arts class were tested; West Liberty (USA) and Royal Melbourne (Australia) Institute of Technology each contributed test data from one chemistry class. Numbers of students tested in each class were 56, 33, 44 and 36, respectively. All students were tested during the twelfth week of the semester.

Split half reliability estimated with chemistry students at Riverina College was .61 (this was considered acceptable for a diagnostic instrument).

Findings

Professor choices of alternatives for the 20 items were considered the "correct" responses. The percentage of students in each of the four groups which selected the correct response for each item was presented and demonstrated that student preferences "differed considerably" from their instructors on almost every item. Mean percentages of correct responses over all items for the four groups were similar (range: 39.4% to 47.7%). Chi-squared statistics were computed for each group for each alternative of each item on the test. Pairwise comparison of college chemistry groups showed significant differences ($p < .05$) in only 25% of the distribution of preferences, indicating that chemistry students in three different college settings are "fairly similar in their preferences for ways of expressing ideas in chemistry on the basis of the test used in this study."

Comparison of chemistry and nonchemistry students at the same institution showed significant differences in choices of alternatives in 45% of the items.

Interpretations

The authors suggested that the difference between language preference of instructors and freshman students suggests that instructors "may become so used to expressing ideas in particular language patterns that they do not recognize that this usage may present a problem to the neonate chemists." They speculated that, if instructors are not sensitive to language, they may tip off the correct answers in multiple choice questions because they phrase the right answer in special ways. A "major implication" of the study was that freshman chemistry students may have to learn appropriate ways of talking and writing about chemistry as they learn the subject matter.

ABSTRACTOR'S ANALYSIS

The authors stated in numerous places the tentative nature of this work. We should believe them. There was no "experiment"; rather, the authors devised an instrument which was then tried out on some available groups. Was there a rationale for selecting the particular groups tested? Were there any differences in the backgrounds of Australian and American students that might have affected the results in any way (for example, number of science classes taken prior to this work, different language patterns in pre-college textbooks, etc.)?

The faculty tested in this work were able to discriminate one preferred way of phrasing a formal chemistry exam response from three (unspecified) alternatives in 80% agreement with one another. Students were generally less successful at selecting the same alternative as faculty. Should we be surprised at this?

The types of items in the language preference test appear to be fundamentally different from one another. Being able to discriminate a system from the model developed to describe the system is a part of the reigning science paradigm. On the other hand, use of the passive

voice does not require new understanding on the part of the students. Specific phraseology only becomes important if students are unable to comprehend what the phrase means or if subtle implications are intended by a specific choice of wording. Yet, all three of these types of items were treated in the language preference test as if they were equivalent.

The items themselves were not well specified. How many of each of the three main types were in the final test? What was the pattern of responses for faculty/student groups within each type of item? Why were these three types chosen? What other types might have been included?

Details of test administration were lacking. Particularly with respect to students, it is unclear what instructions were given to them. Were they to choose the alternative which they preferred or the alternative they believed the professor wanted? Was "preference" linked to "correctness" of response?

We learned very little about the language preferences of students except that they differed from faculty. What were the incorrect responses which students preferred? For example, did they prefer active to passive voice? Was there a pattern? How did the authors generate the incorrect responses? We were told that the correct response had to have three characteristics. What were the characteristics of the incorrect responses? Did they differ from the correct response in one characteristic? Two? There seemed to be no framework within which the incorrect alternative responses were generated.

The results reported were for group means only, with one overall chi-squared statistic. Other chi-squared statistics were calculated but not reported. The statistical work provided only broad generalized conclusions of likeness or difference without any means of determining the strength of the relationships, their potential cause, consequences, etc.

Let's suppose for the moment that a study such as this was able to demonstrate unequivocally that chemistry faculty do prefer certain language patterns in formal test situations and that students do not share this preference. The question then becomes: what difference does that make? Is there any evidence that the dissimilarity causes any difficulties in student learning? Do students who share faculty language preferences learn chemistry better? Are they happier learning chemistry than their peers who use chemical language dissimilar to faculty? At what point in a student's college chemistry career does it become important for him/her to be able to speak like a chemist? Do faculty consider it important that students learn to prefer language like their own in the context of the chemistry examination room or laboratory or other context? If so, how could they teach this skill? Is it directly teachable? In other words, what are the educational consequences of such language dissimilarities? Future work in investigating language preferences of chemistry faculty/students should couple attempts at identifying language preference differences with identifying possible consequences of those differences.

REFERENCES

- Heath, R. W. "Curriculum, Cognition and Educational Measurement." Educational and Psychological Measurement, 24: 248-253, 1964.
- Schoenfeld, R. "The Chemist's English." In Proceedings of the Royal Australian Chemical Institute, 1974.

Yager, Robert E. and Ronald J. Bonnstetter. "Student Perceptions of Science Teachers, Classes and Course Content." School Science and Mathematics, 84 (5): 406-414, 1984.

Descriptors--*Educational Assessment; Educational Research; Elementary Secondary Education; *Science Education; *Science Instruction; *Science Programs; Science Teachers; *Student Attitudes; *Surveys

Expanded abstract and analysis prepared especially for I.S.E. by Eugene L. Chiappetta, University of Houston-University Park, Houston, Texas.

Purpose

In 1982, the authors conducted a follow-up study to the 1978 Third Assessment of Science, part of the National Assessment of Educational Progress (NAEP). Both studies included survey questions which dealt with the affective domain as well as the cognitive domain. It was the affective domain that was the focus of this study, in particular what students and young adults thought of science teachers, science classes, and science subject matter.

Rationale

In the 1970s, educators were voicing discontent about the status of science education because of decline in achievement and interest in science courses. Funding had also decreased during this period. But in the early 1980s a major change occurred, and funding and attention toward science education increased. It was the intent of this study to analyze this period of change between 1977 and 1982.

Research Design and Procedure

The present study was similar to the NAEP study which took place in 1977. Like the NAEP study, it surveyed 9-year-olds, 13-year-olds, 17-year-olds, and young adults. However, unlike the national study, this investigation surveyed people in the State of Iowa.

A pilot study was first conducted in a local Iowa school with class groups at the three age levels to validate a modification of the NAEP instrument. The follow-up study used only 13 items from the NAEP instrument, which are in the affective domain. The pilot study provided information on test administration and scoring these items.

The actual survey took place in the 15 educational service areas throughout the State. It was coordinated by the science consultants in the service areas. Seven consultants randomly selected classes at three age groups (9, 13, and 17 year olds) and, with the help of school counselors, identified adult samples in their communities. This procedure resulted in data from 700 individuals.

Findings

The results of this study give a comparison of survey data that pertain to science teachers, science classes, and science course content for four age groups at two time periods. They indicate that there is virtually no difference between the perceptions of people from four age groups between the 1977 NAEP sample and the 1982 Iowa sample. These results are summarized as follows:

1. Approximately one-half of the elementary teachers admit to not knowing the answers to students' questions, while this drops to one-fifth for the junior high teachers and one-sixth for the senior and adult science teachers.
2. One-third of the elementary students, three-fourths of the junior-high, and two-thirds of the adults feel that their science teachers really liked science.

3. Two-thirds of the nine-year-olds' science teachers, one-half of the junior high science teachers, less than one-half of the senior high teachers, and one-third of the adult science teachers make science exciting for their students.
4. Science classes are interesting to 86 percent of the nine-year-olds, 52 percent of the junior high, 43 percent of the senior high, and 17 percent of the adult students.
5. Science classes make 58 percent of the nine-year-olds feel successful, 36 percent of the junior high, 27 percent of the senior high, and 19 percent of the adults.
6. Approximately 70 percent of the nine-year-olds, 75 percent of the junior high, 80 percent of the senior high, and 40 percent of the adults believe that the science course content they are studying is useful to them.
7. Approximately 90 percent of the nine-year-olds, 75 percent of the junior high, 65 percent of the senior high, and 23 percent of the adults believe that science course content will be useful to them in the future.

Interpretations

The 1982 Iowa follow-up study shows similar results to the 1977 NAEP survey. Secondary science teachers are perceived as knowing the answers and seldom admitting their ignorance about science in their teaching situation. Only a third of the elementary teachers communicate a liking of science to their students; this feeling increases to about three fourths for secondary teachers. Even when admitting to knowing answers, and with few communicating a real liking of science, elementary teachers are most successful with making science exciting for their students. Fewer than half of the secondary teachers are able to make science learning exciting. Elementary teachers are much more successful in encouraging their students to share ideas and experiences. Two thirds of the elementary teachers

are successful, where only 40 percent of the secondary teachers are successful in encouraging sharing of ideas. Science classes are reported to be more fun and interesting for elementary students, but this drops off the longer students stay in school. In addition, students feel more uncomfortable and unsuccessful the longer they continue to enroll in science classes.

The elementary school students believe that science course content is useful to them and will become more so as time goes on. Junior and senior high students generally believe that science content is and will always be important in their lives. About half of the adults believe science subject matter is important, but that it will be less important in their future lives.

ABSTRACTOR'S ANALYSIS

This research report urges teachers and researchers to focus on important outcomes of the science education curriculum, which are in the affective domain. What students feel and believe about science course experiences need immediate attention. The authors urge us to consider seriously the perceptions of students who have enrolled in science courses, and suggest that these data will cause us to recommend a new curriculum for precollege science teaching, especially for the majority of students.

Relationship to other studies. This study clearly fits into investigations related to the affective domain, but it does not build on a line of research in any particular area of this domain. For example, it does not build on the substantial number of studies by science educators who have addressed the curriculum under the heading of scientific literacy. Nor does it build upon the studies related to classroom climate, teacher behavior, and attitude toward science.

Validity. To what extent does this investigation present a true picture of what students believe about their science course experiences? Few would question these results which indicate a decline in interest

toward science, feelings of success in science courses, and beliefs in the utility of science content in everyday life, the longer people take science courses. The data from the Iowa follow-up study are very similar to the national data of the NAEP. Survey data are often validated by comparison with outside criteria such as a national survey.

Research design. The research design is adequate for the follow-up survey in the State of Iowa. Data were collected from seven out of fifteen of the education service areas throughout the State, therefore information was gathered from about half of the educational service areas. One class was selected at random at each age group from each of the areas. This resulted in 100 completed questionnaires from each area with 700 respondents in total. Thus, approximately 175 subjects responded in each of the age groups: 9, 13, and 17 year olds, and adults.

Written report. Most of the article was given to a description of the composition of the questionnaire and the results. This format provides the reader with a clear picture of the nature of the survey. Too often research reports are difficult to understand because most of the report is given to the presentation of empirical data. Nevertheless, the authors let the reader down at the end where they devote only two paragraphs to presenting their recommendations.

The recommendations of this study are too general. The authors urge science educators to make science curricula more relevant by stressing personal needs, resolving major societal issues and problems, and considering careers in science and technology. They further call for renewed attention to the basic ingredients of science, to scientific literacy, to a local/community base, and to the many applications of science.

How can this be accomplished? Science teachers are complaining that their courses are too full and that they cannot put any more into them. In order for recommendations to be useful they must be feasible. If the authors are suggesting a new approach to science teaching from an analysis of these data, they need to be much more specific. Tell us how the present approach, which presumably emphasizes the basic ingredients of science, can be modified to incorporate many of the ideas mentioned and still be feasible to implement.

Suggestions for future research. Future research and the recommendations which emanate from them must be very specific as to the type of science course reforms that should be undertaken, and for whom. The following must be specified:

1. For whom is the "new science curriculum" intended, elementary school pupils, middle school students, or high school students?
2. Is the "new science curriculum" intended for the large percentage of nonscience majors or the small percentage of science majors or both?
3. Roughly what balance should be devoted in the "new science curriculum" to the following:

Basic science content.

Development of thinking skills.

Technological applications of science.

Discussion of moral and ethical issues of science and technology

Clarification of personal values.

Careers in science.

Meaningful dialogue among the science teaching community must focus on these questions and make clear from the start the type of program under discussion. For example, a science/technology/society curriculum may be a reasonable approach for the majority of students enrolled in science courses in grades 5-10 where general science, life science, earth science, physical science, and biology are normally taught. Whereas this type of curriculum might be inappropriate for elementary school science, and high school chemistry and physics which are usually intended for science majors.

The decline in students' positive perceptions about their science teachers and classes is well documented. It begins in the upper elementary school. The data are clear. Why don't researchers thoroughly investigate the reasons for the decline at this point in our educational system?

Lawrenz, F. P. and W. W. Welch. "Student Perceptions of Science Classes Taught by Males and Females." Journal of Research in Science Teaching, 20 (7): 655-662, 1983.

Descriptors--*Classroom Environment; *Females; Performance Factors; Science Education; *Science Instruction; Science Teachers; Secondary Education; *Secondary School Science; Sex; *Sex Differences; *Student Attitudes

Expanded abstract and analysis prepared especially for I.S.E. by John Edwards, James Cook University of North Queensland.

Purpose

To investigate the effect of teacher gender on how secondary students perceived their science classroom learning environments.

Rationale

While at 13 years of age boys and girls are quite similar in science learning, as revealed by analysis of National Assessment results, this similarity does not exist for 17 year olds. One possible cause for this change is socio-cultural pressure. In this study, it is suggested that the perceived science classroom learning environment may contribute to this pressure through differences in classes taught by male and female science teachers.

Research Design and Procedure

The researchers used a ten scale version of the Learning Environment Inventory (LEI) to assess the socio-cultural aspects of the science classroom learning environment. Data were collected from a stratified random sample of science teachers in 14 states. Each state was stratified by city size and representative proportions of secondary schools within each level were randomly selected. Once a school was

selected, a physics, chemistry, biology or junior high school teacher at that school was randomly selected and that teacher then randomly selected one class to complete the LEI. The mean scores for each class on the LEI scales were used as the unit of analysis. Data were obtained from 331 science classrooms, including 110 junior high, 80 biology, 99 chemistry, and 42 physics classes. Approximately 80 percent of the teachers were male and the majority of the female teachers taught junior high or biology classes.

Findings

A two-way multivariate analysis of variance revealed that both teacher gender and type of course were significant factors in accounting for student perceptions of the learning environment while there was no significant interaction. Six of the ten LEI scales showed a significant effect ($p < 0.10$) for teacher gender. The students perceived classes taught by females as more diverse, goal directed, and formal than classes taught by males. Students also viewed classes taught by females as having significantly more friction and instances of teacher favoritism than classes taught by males. Classes taught by males were perceived as more difficult. Comparisons of the average mean obtained scores with a hypothetical neutral score were also used leading to statements of perceived characteristics of science classrooms.

Interpretations

The researchers drew a number of implications from their data in a lengthy discussion session. They suggested that the differences reported may help to explain why girls move away from science. In concluding their paper, the authors provided a number of cautionary statements with respect to their data, their analysis, and their perceived implications. Rather than attempt to list these, they are dealt with, in part, in the analysis that follows.

ABTRACTOR'S ANALYSIS

The study by Lawrenz and Welch deals with one aspect of an area of increasing focus in science education: gender factors. The MANOVA design is a good conceptualization for such a study and the size of the sample suggests the possibility of widely generalizable findings. Three major areas of concern arise in the paper:

- (i) awareness of other studies in the area;
- (ii) the statistical analysis; and
- (iii) internal consistency in the paper.

(i) Awareness of other studies in the area: A study such as this fits into a broader context of gender factor studies in education. The paper contains many conjectures which bear little, if any, relation to the data obtained in the study. To find that these conjectures reveal little awareness of the considerable current literature is disturbing. This can be exemplified by taking three conjectures from the paper and setting them in the broader context mentioned above.

First conjecture

"It seems likely that students make decisions about future science courses or careers based on their perceptions of their present science classes."
(Lawrenz & Welch, p. 656)

Delamont (1980), Kelly (1981), Spender (1982) and Stockard et al (1980) reported discriminatory practices in schools, particularly at the primary level which serve to steer girls into subjects other than mathematics and science. Delamont (1980) concluded that girls at the primary level receive up to eight times less instruction in problem solving than boys. Kelly (1981) reported that girls suffer from discrimination in treatment by counsellors, some teachers, and boys in their classes, which effectively closes off the option of science. Spender (1982) demonstrated that two-thirds of classroom time is spent with the boys in a class. Stockard et al (1980) pointed out the

deficiency in school text books in providing role models for girls beyond stereotypes of women in the home. Connell et al (1982), Rowlands (1980), and Wood (1976) indicated that subject choice must be seen within the wider context of class and gender power structures. Keys and Ormerod (1976) and Ormerod (1975) indicated that social advantages outweigh other factors in subject choice. Ormerod (1975) highlighted the difference between subject preference which correlated with liking for teachers, and actual subject choice which showed no relationship with attitude towards teachers.

Second conjecture

"It is certainly not unlikely that a girl contemplating a future career in a science oriented field would begin to view it as somewhat deviant behavior." (Lawrenz & Welch, p. 659)

The failure of Lawrenz and Welch to distinguish between the physical and biological sciences reduces the credibility of this statement. Fox and Denham (1974), Hansen and Neujahr (1974) and Koelsche (1965) observed equal male-female interest in biological sciences. Gardner (1975), Keeves and Read (1974), and Kelly (1976) reported similar proportions of girls and boys enrolled in biological science courses. The pattern of enrollment of the sexes has been paralleled by patterns of achievement. Comber and Keeves (1973) indicated that any advantage boys might have over girls in achievement test scores tends to be small in the biological sciences. Moreover, evidence from the NAEP studies strongly suggests that differences such as these do not increase with age (NAEP, 1977).

Third conjecture

"So perhaps, as Bossert (1981) suggested for mathematics, girls view success in science classes as 'unfeminine'; an image few teenage girls would relish. This view of science as unfeminine ..."

(Lawrenz & Welch, p. 660)

Fennema and Sherman (1977) and Lantz (1980) found no correlation between 'Mathematics as a Male Domain Scale' and course participation. Boswell and Katz (1978), Brush (1979), Fennema and Sherman (1978, 1977, 1976), Fox (1979), and Sherman (1980) reported that females stereotyped mathematics as a male domain significantly less frequently than did males. This would lead one to doubt the above conjecture.

When conjectures in their paper are viewed against this broader background, Lawrenz & Welch's admission: "It must be pointed out that these are suggestions" (p. 660) is seen in a different light. For those interested in the area of gender and ethnic factors in choice of sciences, Stead's (1983) study using personal construct theory presents an interesting approach.

(ii) Statistical analysis: The authors' acceptance of all ten LEI scales is questionable, given the modest Cronbach alpha figures displayed in Table I (p. 657). There are good reasons to regard scale alpha levels below 0.80 with some caution, and this would challenge at least five of the ten LEI scales. While it may be that Lawrenz & Welch have unreported evidence to confirm the scale properties of the LEI instrument, the information published in this paper raises unresolved doubts about issues of homogeneity and unidimensionality (see Green et al, 1977).

The question should also be raised as to whether it is appropriate to define as 'neutral' a score of 2.5 on a 4-point Likert scale. While it may be reasonable to adopt this purely analytic approach, without supporting data for these scales it is difficult to justify a definition of this kind. The overall means for the ten scales, reported in Table IV, are Males: 2.5, Females: 2.55, which suggests that 2.5 may not in fact be the midpoint in any educationally meaningful sense.

It is always open to a researcher to set aside what are, too often, blindly practiced conventions of statistical analysis - provided there are reasons to defend such a decision. In this respect, it is interesting to note that common practice would have required p values to be set at or below the 5 percent level before claims of significance

could be entertained, but Lawrenz and Welch chose the 10 percent level, thus effectively identifying six (rather than four at .05) LEI scales as 'significant'. The concern this raises deepens when this apparently arbitrary relaxation of the conventional significance rules comes on top of less-than-rigorous decisions regarding the scales themselves. As a result of apparent acceptance of relatively soft statistical decisions, Lawrenz and Welch claim substantive importance for the differences in their data between science classes taught by males and those taught by females. They then go beyond this to discuss the implications of their results in terms of classroom organization and subject choice options for students. In fairness, the last three paragraphs of the paper reveal reservations as to the cogency of Lawrenz and Welch's data. However, it is highly pertinent at this stage to reconsider their paper from a more conventional statistical point of view.

In terms of widely recognized practice, only four of the LEI scales (Formality, Friction, Favoritism and Difficulty) would be counted significant following the analyses based on MANOVA. Of this group, Friction and Difficulty could not be regarded as acceptable in view of their weak levels of internal consistency (alpha's of 0.72 and 0.64 respectively). If one then accepts a conservative decision rule allowing that in any set of ten analyses, one result might be expected to produce significant results on chance alone, by selecting the more favourable we are left with only one scale difference (on Formality) involving a shift of 0.08 scale points - or a little more than one-third of a standard deviation.

It should be made clear that Lawrenz and Welch are under no obligation to select the conservative option at each decision point in this sequence. Provided they can justify their choices, they should follow whatever path best fits their data and the nature of the probable consequences. In view of their failure to offer any justification for the relaxation of conventional rules, it is difficult to take the study's results as seriously as might otherwise have been done.

(iii) Internal consistency: There is a tenuous relationship in the paper between the introductory section, the data, and the discussion. For example, Haertel et al (1981) are cited to highlight differences between the sexes during adolescence in science related attitudes, behaviors and aspirations. Yet in this study no record was reported of the male to female ratio in the classes. This is particularly important when one refers to the studies reported earlier in this analysis. Chemistry and physics classes commonly have a higher proportion of male students than do biology classes. One could also expect that the impact of peer relationships would be greater in activity-centered classes, with the impact of the teacher being more dominant in teacher-centered classes.

While the reported data deal with students' attitudes towards science instruction in relation to sex of teacher, the discussion diverges into conjectures involving areas such as sex-roles of students and academic background of teachers. The problems of such divergences are compounded when seen in the broader context outlined earlier in this analysis.

In summary, this paper is a disappointing contribution to the study of gender factors in science education. Those interested in the area are encouraged to set the data from this study in the broader contexts proposed in this analysis.

REFERENCES

- Boswell, S. L. and P. A. Katz. Nice Girls Don't Study Mathematics. ERIC No. 188 888, 1978.
- Brush, L. R. Why Women Avoid the Study of Mathematics: A Longitudinal Study. ERIC No. ED 187 887.
- Comber, L. C. and J. P. Keeves. Science Education in Nineteen Countries. Stockholm: Almqvist & Wiksell, 1973.
- Connell, R. W., D. J. Ashenden, S. Kessler, and G. W. Gowsett. Making the Difference. Sydney: Allen and Unwin, 1982.
- Delamont, S. Sex Roles and the School. London: Methuen, 1980.

- Fennema, E. and J. Sherman. Fennema-Sherman Mathematics Attitude Scales: Instruments Designed to Measure Attitudes Toward the Learning of Mathematics by Females and Males. Catalog of Selected Documents in Psychology, 6 (1): 31, 1976.
- Fennema, E. and J. Sherman. "Sex Related Differences in Mathematics Achievement, Spacial, Visualization and Affective Factors." American Education Research Journal, 14: 51-71, 1977.
- Fennema, E. and J. Sherman. "Sex-Related Differences in Mathematics Achievement and Related Factors: A Further Study." Journal for Research in Mathematics Education, 9: 189-203, 1978.
- Fox, L. H. Women and Mathematics: The Import of Early Intervention Programs Upon Course Taking and Attitudes in High School. ERIC No. ED 188 886, 1979.
- Fox, L. H. and S. A. Denham. "Value and Career Interest of Mathematically and Scientifically Inclined Youth." In J. C. Stanley, D. P. Keating, and L. H. Fox (eds.) Mathematical Talent: Discovery, Description, and Development. Baltimore: Johns Hopkins, 1974.
- Gardner, P. L. "Sex Differences in Achievement Attitudes and Personality of Science Students: A Review." Science Education: Research, 1974, 2: 1-41, 1975.
- Green, S. B., R. W. Lissitz, and S. A. Mulaik. "Limitations of Coefficient Alpha as an Index of Test Unidimensionality." Educational and Psychological Measurement, 37: 827-838, 1977.
- Hansen, R. and Neujhar. "Career Development of Males and Females Gifted in Science." Journal of Educational Research, 68: 43-45, 1974.
- Keeves, J. P. and A. D. Read. Sex Differences in Preparing for Scientific Occupations. IEA (Australia) Report 1974: 6 ACER, 1974.
- Kelly, A. "Women in Science: A Bibliographic Review." Durham Research Review, 67: 1092, 1976.
- Kelly, A. The Missing Half: Girls and Science Education. Manchester: University Press, 1981.
- Keys, W. and M. B. Ormerod. "Some Factors Affecting Pupils' Subject Preference." Durham Research Review, 7: 36, 1976.
- Koelsche, C. L. "Characteristics of Potential Scientists." Science Education, 49: 72-79, 1965.
- Lantz, A. Determining the Importance of Factors Influencing the Election of Mathematics Courses. Final Report, Denver Research Institute, January 1980.

- National Assessment of Educational Progress (NAEP). Science Technical Report: Summary Volume, Denver: Education Commission of the States, 1977.
- Ormerod, M. B. "Subject Preference and Choice in Co-education and Single Sex Secondary Schools." British Journal of Educational Psychology, 45: 257-267, 1975.
- Rowlands, R. G. "Are Girls a Disadvantaged Group?" Australian Journal of Education, 20: 21-37, 1980.
- Sherman, J. Women and Mathematics: Summary of Research from 1977 to 1979. ERIC No. ED 182 162, 1980.
- Spender, D. Invisible Women: The Schooling Scandal. London: Writers and Researchers, 1982.
- Stead, K. "Insights into Students' Outlooks on Science With Personal Constructs." Research in Science Education, 13: 163-176, 1983.
- Stockard, J., P. Schmuck, K. Kemper, P. Williams, S. Edson, and M. A. Smith (eds.) Sex Equity in Education. New York: Academic Press, 1980.
- Wood, P. "The Myth of Subject Choice." British Journal of Sociology, 29: 130-149, 1976.

RESPONSES TO CRITIQUES

IN RESPONSE TO THE ANALYSIS OF

Yager, Robert E. and Ronald J. Bonnstetter. "Student Perceptions of Science Teachers, Classes and Course Content" by Eugene L. Chiappetta. Investigations in Science Education 12 (4): 47-52, 1986.

Robert E. Yager
The University of Iowa

Perhaps it is the time to suggest major departures from the science programs currently found in schools. The results of the 1978 NAEP Assessment and the 1982 follow-up suggest that the current situation leaves much to be desired in terms of long term and lasting affective outcomes. The declines in achievement measures have been a source for general alarm for well over a decade. The NSF Status Studies of 1977 and 1978 also suggest major failures of traditional curricula and instructional practices. There is no evidence that more of the same will result in any correctives. And yet there seem to be many, including those who are proposing and funding new (?) science education programs in 1985, who want to try again with the same correctives undertaken during the two decades following the launching of the Soviet Sputnik.

Chiappetta starts with the assumption that corrective actions must be add-ons, i.e., slight alterations of the existing programs and teaching practices. Perhaps new programs and new teaching are needed to resolve the problems and to reverse the negative results of school science experiences for most students.

A new science program may start with societal issues, move to a consideration of technology upon the current human existence, and advance to the more abstract principles of science for as many as possible. Currently we commonly begin with science, with the belief and assertion that students must first know science before they can apply it and before they can consider the societal issues involved. If this is accurate, most students are lost (perhaps 90%) before any applications and any social issues are considered.

Perhaps it is more logical and more meaningful for most students to begin with issues and problems as viewed by humans in their environment/culture. In the world culture at the end of the Twentieth Century, science and technology are widely recognized as the dominant forces. They influence every aspect of our existence, our survival, our careers, our recreation, our lives. If science and technological issues are central to modern society, they must be central in the educational/schooling enterprise.

The current Science/Technology/Society focus - though ill-defined - is recognized as the newest "movement" in U.S. science education. Some would argue that it would be better known as Society/Technology/Science--a reversal of the "S's" to emphasize the human being--his/her society--as a starting point for most. This would lead, in our current society, to a consideration of the value, centrality, and danger of technology in the lives of all. To understand technology and the societal issues a basic knowledge of science becomes obvious--even for those students not able to understand the ideas, concepts, principles, theories of science thoroughly.

Such a focus would help students see the value in science, i.e., knowing its basic content as opposed to teaching about its value as religion--faith of its importance.

Students should be involved in the resolution of real issues and problems in their homes, schools, communities, counties, states. Involved students become learners. Knowledge of science becomes an outcome--not a starting point.

To search for a perfect science program--one with just the "right" degree of content per se, career information, focus on societal problems, and other goal areas may be a problem in and of itself. The curriculum structure in science should evolve; perhaps it can be described after an experience of a year or more for a student or group of students.

In many respects instruction may be more of an issue than curriculum structure. This suggests the importance of the teacher's grasp of the meaning of science, the current situation, student interests and abilities. The particular format for the science curriculum may be

likened to concern for the make of a car or its structure - its appearance. But what is the function of the car? Where is it going? Why? For what point? And, is the best one for one individual in one particular place with one particular goal the same for all other situations? Maybe balance for one teacher, one student, one class, one school, or one state will be different than that for others.

IN RESPONSE TO THE ANALYSIS OF

Lawrenz, Frances and Wayne Welch. "Student Perceptions of Science Classes Taught by Males and Females." by John Edwards. Investigations in Science Education 12 (4): 53-61, 1986.

Frances Lawrenz
Arizona State University

Wayne W. Welch
University of Minnesota

We are puzzled as to the purpose of Edwards' scathing and patronizing criticism of our paper. Apparently, he has misunderstood a number of key points and finds it difficult, if not impossible, to believe the findings of our study. We would like to respond to each of his three areas of "concern."

First, as to awareness of other studies in the area, all of our "conjectures" were amply supported by references to other research in the area. To support his critique Edwards takes three sentences out of context and mentions research we did not reference that pertains to the ideas expressed in them. This is almost always possible in a study that is not a comprehensive research review. The first "conjecture" Edwards selected was from our introduction and is peripheral to the intent of the study which was to determine if perceptions of the classroom environment might be related to the changes in girls' interest in science reported by Haertel et. al. (1981). Certainly we never meant to imply, and we don't believe we did imply, that perceived science classroom environments are the sole reasons or even causes at all for future career decisions. As we carefully state in our discussion section, we were reporting exploratory research to help find variables that might be related to the girls' change in interest in science.

The second and third sentences Edwards selected relate to whether or not girls view science as a masculine endeavor. Edwards points out that findings for the biological sciences are different than for the physical sciences. That is true but our data showed no interactions

of the perceptions of students from biology, chemistry or physics classes with teacher gender. In other words, students in all three types of classes viewed male and female teachers in the same way. We could not, therefore, consider biology separately. Edwards also suggests that data on mathematics may not support the idea that science is masculine. Certainly that is possible, but the research we cited holds the contrary opinion. In all, we believe we proposed thoughtful interpretations of our data and provided adequate references to support them.

Edwards' second area of concern was our statistical analyses. The LEI has been used extensively, and the reliability and validity of the scales have been carefully documented so we believe the concern over the scale alpha levels is unwarranted. The 2.5 neutral score was assumed because it was purely analytical and no "educationally meaningful" midpoints have been established. Our findings were based on the results of a two step statistical procedure. We initially conducted a MANOVA combining all of the LEI scales to ascertain if any differences existed between perceptions of classes taught by males and females. This analysis was significant at the .00 level; certainly a 'significant' finding by any standards. Clearly differences did exist. Once this strong evidence was obtained, we examined the univariate results to determine which scales contributed the most to this multivariate effect. A p level of .1 was chosen for these analyses because this study was exploratory and to err in favor of possible variables was more appropriate.

Edwards' third area of concern, lack of information on the number and perceptions of boys and girls in the classes, was well taken. We also suggest that future research should examine these questions. In this study, however, we used class mean scores as the basis for the analysis and therefore were unable to examine individual gender differences.

We conducted our research as an exploration to try to determine if the perceived classroom psychosocial environment could be related to changes in girls' interest in science. We presented original data and made no attempt to conduct an exhaustive review of existing research. We used a somewhat more liberal significance level after the MANOVA had

shown significance and an extended discussion section because the study was exploratory. We were not trying to eliminate possible variables but to find them. Edwards' criticisms of our study are based on a different set of goals and as such produced a much more critical review than we believe was justified.